「発言小町投稿データ」CSV ファイル仕様書

読売新聞東京本社メディア局事業部

1.「発言小町投稿データ」ファイル仕様説明書 Version1.0

修正履歴

2020.10.8 Version1.0 作成

2022.1.21 Version2.0 2021 年に実施したリニューアルを受けて改版

- ①文字コードは UTF-8。(2019、2020 年版は Shift-JIS)
- ②サイトがレスポンシブ化して SP と PC の区別がなくなったため、「投稿デバイス」は「(PC) かそれ以外) PC または other」から「PC (SP) またはスゴ得」に
- ③ユーザーID 10 桁から 16 桁に変更(21 年中のリニューアルのため、同年データには 10 桁と 16 桁が混在)
- ④顔アイコンを、独自アイコンから UTF-8 の顔文字へ変更 (一部、独自アイコン。表参照)
 - ⑤お気に入り登録が会員のみの機能となった。非会員の場合は0。

2. ファイルについて

- ・発言小町(https://komachi.yomiuri.co.jp/)に掲載されたトピを、1 か月につき 1 CSV ファイルにして 作成しています。
- ・文字コードは UTF-8 を使用しています。
- ·CSV ファイルの命名規則は以下の通りです

【西暦 4 桁】 【月 2 桁】.csv

- ・公開済みのトピ・レスのみ出力しております。削除された物は CSV に出力されていません。
- ・トピの書き込み日時に従って、月ごとの CSV ファイルを作成しました。
- ・ファイル作成時点で公開されていたトピ・レスのみ出力しております。出力後も公開中のトピには、ファイル作成以降にレスがつく可能性があります。
- ・これポチ投票数はファイル作成時の値です。お気に入り登録数もファイル作成当時の値です。

3. データ形式

	データ	形式	最大長 (byte)	データ内容・備考
1	トピ ID	整数	7	トピ ID(2020 年 12 月以前分は 6 桁)
2	トピ投稿日時	日付		トピの投稿日時 YYYY/MM/DD hh:mm:ss
3	トピの書き込みが PC かそれ以 外か	文字列	5	投稿デバイス、PC(SP) or スゴ特
4	長いジャンル名	文字列	30	長いジャンル名
5	短いジャンル名	文字列	6	短いジャンル名
6	トピ名	文字列	128	トピタイトル
7	レス数	整数		レス数
8	トピ主レス数	整数		トピ主のレス数
9	ユーザーID	文字列	16	ユーザーID
10	トピ主名	文字列	64	ハンドルネーム
11	トピ主アイコン	整数	6	UTF-8 の顔文字。Id(int)で出力可能。
12	トピ本文	文字列	5000	トピ本文
13	これポチ面白いの投票数	整数		面白いの投票数
14	これポチビックリの投票数	整数		びっくりの投票数
15	これポチ涙ぽろりの投票数	整数		涙ぽろりの投票数
16	これポチエールの投票数	整数		エールの投票数
17	これポチなるほどの投票数	整数		なるほどの投票数
18	レスタイトル	文字列	128	レスタイトル
19	レス ID	整数	4	トピ内でふられたレスの番号。レス番号 は、ゴミ箱分の数字が歯抜けになります。 連続していません。
20	レス投稿日時	日付		レスの投稿日時 YYYY/MM/DD hh:mm:ss
21	レスの書き込みが PC かそれ以 外か	文字列	5	投稿デバイス、PC(SP) or スゴ特
22	レスのユーザー名	文字列	64	ハンドルネーム
23	レスのユーザーのアイコン	文字列	6	顔アイコン No、データ内容は下記参照
24	レスのユーザーID	整数	16	ユーザーID
25	レス本文	文字列	5000	レス本文
26	親トピ投稿日時	日付		レスに対する親トピの投稿日時 YYYY/MM/DD hh:mm:ss

※注1)トピ主とレスのユーザーIDは、各トピの中でのみユニークです。別のトピとはユニークとはなっておらず重複する可能性があります。

4. データ内容について

4-1 ジャンル名

短いジャンル名と長いジャンル名は下記のように対応しています。

短いジャンル名	長いジャンル名
話題	生活・身近な話題
恋愛	恋愛・結婚
夫婦	夫婦関係・離婚
子供	妊娠・出産・育児
仕事	キャリア・職場
ひと	家族・友人・人間関係
ヘルス	心や体の悩み
美	美容・ファッション

4-2-1 顔アイコン

UTF-8の Id(int)で出力可能なアイコンを使用しています。

ID	name	html コード	見た目
1	grinning face	😀	(1)
2	winking face	😉	(3)
3	smiling face with heart-eyes	😍	(1)
4	fearful face	😨	®
5	crying face	😢	(E)
6	loudly crying face	😭	
7	persevering face	😣	8
8	expressionless face	😑	=
9	angry face	😠	(2)
10	squinting face with tongue	😝	(#)

11	unamused face	😒	3
12	pouting face	😡	&
13	dog face	🐶	a
14	cat face	🐱	E
15	baby chick	🐤	\mathfrak{D}
16	penguin	🐧	₫
17	horse face	🐴	(A)
18	pig face	🐷	<u>e</u>
19	sun	☀	*
20	cloud	☁	•
21	umbrella	☂	†
22	snowman without snow	⛄	•
23	high voltage	⚡	4
24	red heart	❤	\Diamond
25	broken heart	💔	♡
26	musical notes	🎶	5
27	hot springs	♨	55
28	kiss mark	💋	0
29	sparkles	✨	\$
30	light bulb	💡	\$

31	anger symbol	💢	ĭĬ
32	ZZZ	💤	£\$2
33	fork and knife	🍴	99
34	wrapped gift	🎁	1
35	passenger ship	🛳	
36	lipstick	💄	A
37	money bag	💰	⑤
38	ring	💍	8
39	birthday cake	🎂	
40	slightly smiling face	🙂	
		旧絵文字	
41	purple circle	kao040	•
		と同じ画像	
42	汗かいた笑顔	😅	
43	思考中の顔	🤔	9
44	笑い転げる顔	🤣	3
45	マスク顔	😷	a
46	口が開いた顔	😮	\odot
47	恐怖で叫ぶ顔	😱	©
48	鼻息の荒い顔	😤	@

		 旧絵文字	
999	くらげっと	kao999	9
		と同じ画像	

4-2-2 2019、2020年版の旧顔アイコンについて

以下の顔アイコン表と発言小町ウェブページとの対照表は、2019、2020 年版で使用していた内容です。 2021 年版以降は使用していません。

顔アイコン一覧表

	alt 値	顔アイコン No
	嬉しい	kao000
:	わーい	kao001
•	ウィンク	kao002
••	目がハート	kao003
(19	がっかり	kao004
3	涙	kao005
***	号泣	kao006
x	がまん	kao007
•	ボケー	kao008
1	むっ	kao009
*	あっかんべー	kao010
•	ふてくされ	kao011
***	激怒	kao012
	犬	kao013
	猫	kao014
%	ひよこ	kao015
<u>♣</u>	ペンギン	kao016
A	ウマ	kao017
(ii)	ブタ	kao018
*	晴れ	kao019

4	曇り	kao020
÷	雨	kao021
8	雪	kao022
<i>4</i>	雷	kao023
v	ハート	kao024
90	失恋	kao025
ð	るんるん	kao026
₩	温泉	kao027
*	キスマーク	kao028
*	ぴかぴか	kao029
≱	ひらめき	kao030
*	むかっ	kao031
Zz	眠い	kao032
41	レストラン	kao033
m	プレゼント	kao034
△	リゾート	kao035
<u>△</u>	化粧	kao036
•	お金	kao037
•	指輪	kao038
***	ケーキ	kao039
•	使用しない	kao040
無	くらげっと	kao999

